Toward an Understanding of Data Literacy

Hammad Rauf Khan[1], Jeonghyun Kim, and Hsia-Ching Chang

[1] University of North Texas, Denton TX 76203, USA
hammadkhan@my.unt.edu

Abstract. As the interest in data grows, much attention has been paid to data literacy, and multiple perspectives and understandings to define data literacy have emerged from varying conceptual contexts. However, there remains a lack of agreement regarding the scope of data literacy across disciplines. This study attempts to define data literacy holistically through a meta-synthesis approach. The study found three distinct themes for data literacy: as skills required for data-driven decision making, as activities for research data services, and a set of practices for data lifecycle.

Keywords: Data, Big Data, Data Literacy, Data-Driven Decision Making, Research Data Service, Data Lifecycle

1 Introduction

Technological advancement has resulted in large quantities of data being generated by individuals, groups, and organizations. The various sources and multiple formats have produced such a massive quantity of data that can be referred to as "big data." The ever-increasing growth of big data has impacted every aspect of our modern society, including business, marketing, government agencies, health care, academic institutions, and research in almost every discipline. Big data has begun to challenge the academic community's capacity to rapidly learn and implement the concepts, techniques, and tools necessary to be considered data-literate, however, the disciplines that promote data literacy have not reached a consensus on the scope and definition of data literacy. This paper will explore how data literacy is defined and in what context it is presented.

2 Methodology: A Meta-Synthesis Approach

This study adopted a meta-synthesis approach, which integrates, compares, and analyzes related studies to form an interpretive synthesis based on themes discovered in the literature (Urquhart, 2010; Togia and Korobili, 2014). To identify relevant articles, a search was conducted using multiple scholarly databases, including Web of Science, Google Scholar, and Library Information Science Technology Abstracts. The primary search terms used to locate relevant articles were any derivative forms of data – i.e., data, big data, research data, etc. These terms were added to a set of secondary search

terms, including literacy, literate, competency, knowledge, and skill. This search resulted in a total of 69 articles.

The analysis involved the following stages: 1) compare and contrast phrases, ideas, concepts, and themes of data literacy represented in the original articles; 2) undertake reciprocal and refutational translations to establish how the themes arising from the included articles were similar or different; 3) synthesize the themes arising from the preceding step.

3 Preliminary Results

The meta-synthesis analysis revealed the following three themes that capture the conception of data literacy. Three main disciplines that contributed to defining data literacy include education, library and information science, and information systems. A few representative definitions are shown in Tables 1, 2, and 3.

3.1 Data Literacy as Skills Required for Data-Driven Decision Making

In early 2000, data literacy was regarded as a synonym of statistical literacy, which refers to the knowledge and skills that enable data users to understand, evaluate, and communicate statistical data in the field of social science and education. This conception was supported by the International Association for Social Science Information Services and Technology and the Association of Public Data Users (Shield, 2004). In recent years, however, the term data literacy has re-emerged as a term of the data revolution discussion, mainly from the perspective of data analytics and data science, which supports data-driven decision making.

Table 1. Data literacy as skills required for data-driven decision making

Carlson et al. (2011)	"…Understanding what data mean, including how to read graphs and charts appropriately, draw correct conclusions from data, and recognize when data are being used in misleading or inappropriate ways."
Johnson (2012)	"…The ability to process, sort, and filter vast quantities of information, which requires knowing how to search, how to filter and process, to produce and synthesize."
Love (2004)	"…The ability to examine multiple measures and multiple levels of data, to consider the research, and to draw sound inferences."
Wolff et al. (2017)	"…Data literate reader has the ability to properly evaluate the evidence that is presented in these scenarios, so that they can make critical judgments on the reliability of the information presented and can better understand how their own contributed data is being utilized and make more informed decisions."

3.2 Data Literacy as Activities Supporting Research Data Services

Since the emergence of e-science and cyberinfrastructure, data literacy has been recognized as one of the critical building blocks in the knowledge base of information professionals involved in supporting research data management and curation (Koltay, 2016).

Additionally, data literacy often has been conceptualized as a guiding principle for activities supporting research data service, which is a campus-wide program that provides the institution community with the expertise, tools, and infrastructure necessary to manage and steward research data. As such programs have been structured on the belief that there is great potential to match existing librarians' expertise in information literacy with support for e-science, data literacy is often viewed as a logical extension of the information literacy concept or as a sub-discipline of information literacy (Schneider, 2013).

Table 2. Data literacy as activities supporting research data services

Koltay (2016)	"…Is closely related to research data services that include research data management."
Prado & Marzal (2013)	"…Enables individuals to access, interpret, critically assess, manage, handle and ethically use data."
Qin & D'Ignazio (2010)	"…Knowledge and skills in collecting, processing, managing, evaluating, and using data for scientific inquiry."
Searle et al. (2015)	Identified as "one of the research data service activities that support researchers in building the skills and knowledge required to manage data well."

3.3 Data Literacy as a Set of Practices for Data Lifecycle

The last definition is a combination of the two definitions discussed so far. This definition of data literacy considers a set activities and practices undertaken in the production, consumption, and management of data through the data lifecycle. Such activities and practices include data creation, data acquisition, data normalization, data analytics, data storage, data disposal, and data use/re-use.

Table 3. Data literacy as a set of practices for data lifecycle

Carlson et al. (2011)	Components of data literacy competency include: "discovery and acquisition; databases and data formats; data conversion and interoperability; data processing and analysis; data visualization and representation; data management & organization; data quality and documentation; metadata and description; cultures of practice; ethics & attribution; data curation and re-use; and data preservation."
Matthews (2016)	"…a set of practical competencies situated in a wider context of personal and social goals, as well as challenges such as subjectivity and bias."
Maybee & Zilinski (2015)	Comprised of "skills related to accessing, managing, communicating, preserving and ethically using data."

Okamoto (2017)	"…The ability to access, critically assess, interpret, manipulate, manage, summarize, handle, present, and ethically use data."

4 Conclusion

The preliminary results of the meta-synthesis reveal that data literacy has multiple facets and is applied with reference to its actions, objective, and context. Confining data literacy to a single discipline or forming multiple literacies under different disciplines with separate terminology only stands to befuddle users. Data literacy is relevant to all disciplines and areas that use data for analysis and decision making and to monitor trends. Koltay (2016) has also emphasized the importance of unifying the similar concepts and terms revolving around data literacy. Terms such as science data literacy and data information literacy are limiting in its reach and do not encompass the broader approach to what data literacy is trying to accomplish through data-driven decision making, research data services, and data lifecycle management. There is a need for flexible and scalable approaches that take into consideration the diversity of data stakeholders from a range of disciplines.

5 Future Research

Moving forward it is important to continue the work toward a universal definition and scope of data literacy. The current meta-synthesis study is not all inclusive and has many limitations. Comprehensive analysis of all available data is impossible, particularly in the confines of one individual study. Therefore, it is important to note that all research on this topic was not incorporated in the culmination of this analysis and that future research should take a more comprehensive approach in defining data literacy.

Selected References

1. Koltay, T. (2016). Data governance, data literacy and the management of data quality. *IFLA Journal, 42*(4), 303-312.
2. Schneider, R. (2013). Research data literacy. *Proceedings of the 1st European Conference on Information Literacy* (pp.134-150).
3. Shield, M. (2004). Information literacy, statistical literacy, data literacy. *IASSIST Quarterly, 28*, 6–11.
4. Togia, A. A., & Korobili, S. (2014). Attitudes towards open access: A meta-synthesis of the empirical literature. *Information Services & Use, 34*(3/4), 221-231.
5. Urquhart, C. (2010). Systematic reviewing, meta-analysis and meta-synthesis for evidence-based library and information science. *Information Research, 15*(3) colis708.